International Journal of Educational
Science and Research (IJESR)
ISSN(P): 2249-6947; ISSN(E): 2249-8052
Vol. 7, Issue 1, Feb 2017, 167-178

THE IMPACT OF VARIATION THEORY FOR SECONDARY SCHOOL STUDENTS ON ACQUIRING PROBABILISTIC KNOWLEDGE

DOI PEI YICK & CHI KEUNG LEUNG

The Education University of Hong Kong

ABSTRACT

This study explored how the teaching of probability under the framework of Ference Marton's Variation Theory affected the learning outcomes of secondary 5 students. After eight probability lessons, the experimental group of 25 secondary 5 students significantly improved their posttest mean score on the overall probabilistic skills by 24% from the pretest mean score under the 90% significance. In contrast, the control group of similar ability (n = 30) unfortunately showed no remarkable progress in terms of their transfer performance. Regarding the elimination of probabilistic misconceptions, the experimental group was more capable of avoiding two of the four measured misconceptions, as contrasted to the control group of similar ability who made no progress on any of the misconception aspects, while the control group of higher ability (n = 36) improved on all four measured aspects.

Variation Theory helped senior secondary students learn probability, especially for those who were at average ability, while the students with higher learning ability may still learn probability well under the traditional teaching approach.

A group of secondary 1 students (n = 32) also wrote the same assessment. The probabilistic knowledge of the secondary 1 student was compared with the initial performance of the secondary 5 students. No significant difference was detected between the secondary 1 and 5 students. This study agreed that the probabilistic skills did not necessarily grow with ages.

KEYWORDS: Variation Theory, Probability, Mathematics Instruction & Transfer Performance

Received: Jan 02, 2017; **Accepted:** Jan 30, 2017**; Published:** Feb 03, 2017**; Paper Id.:** IJESRFEB201720

Original Article

INTRODUCTION

Probability is distinctive not only in its nature, but also its misconceptions students may encounter. Unlike other branches in mathematics, such as trigonometry which can always be proved by geometric constructions, proofs in probability are not always available (Ang & Shahrill, 2014). Probability does not necessarily relate to students' mathematics skills in other domains, as Rakes (2010) draws from his study with more than 1100 high school students that the misconception in probability did not have a casual effect on rational numbers, algebra or geometry misconceptions. Yet, probability can be demanding to students on cognition because it can be differently interpreted based on various meanings, namely (i) intuitive meaning, (ii) classical meaning, (iii) frequentist meaning, (iv) subjective meaning, and (v) mathematical thinking (Batanero and D´ıaz, 2012). The non-algorithmic nature makes probability both unique and demanding to student conception.

The major reason why students find probability arduous is they struggle with different kinds of misconceptions. There are many possible misconceptions of probability students may have in solving probability problems. Li (2000), from the response of over 500 grades 6, 8 and 12 to 83 probability questions, identifies 14

probability misconceptions, namely:

"(1) subjective judgments, (2) example-based interpretations for possible and impossible, (3) possible means certain, (4) chance cannot be measured mathematically, (5) equiprobability, (6) outcome approach, (7) one trial is unrelated to other trials, (8) interpreting chance by data matching or word matching, (9) increasing repetitions is not better for predicting, (10) positive and negative recency, (11) used own methods in chance comparison, (12) taking different order as the same, (13) misuse or extend conditions inappropriately, and (14) used own methods of chance calculation" (p.93).

Similarly, Ang and Shahrill (2014) categorize four types of misconceptions on probability, namely representativeness, equiprobability bias, beliefs and human control, based on the participation of 177 Year 10 and 11 students in their research. Representativeness and equiprobability bias appear to be common misconception when students handle probability. Various misconceptions that students have may hinder students solve probability problems.

Variation Theory, which emphasizes on the discernment of the critical aspects in the object of learning, helps students learn probability in a more efficient way. Originated by Ference Marton, Variation Theory allows students to learn content subject in a layered approach through encrusting the critical aspects of the object of learning while remaining the other aspects constant (Lam, 2013). The object of learning is the "knowledge" of a subject that teachers expect students to understand in the lessons and in a long run it is the "capability or attitude" that students acquire (Ekawati& Lin, 2014). For student learning a particular object of learning, teachers should provide students with different patterns of variation (Cheng, 2016). The theory advocates four patterns of variation: contrast (simultaneously varying an attribute and its related, yet mutually exclusive, attributes), separation (varying an attribute in different values), generalization (fixing an attribute while varying other attributes), and fusion (including two of the patterns at the same time) (Olteanu & Olteanu, 2013).If the teaching of probability, which is full of abstraction to students, is presented by teachers and taught to students based on the framework of Variation Theory, incorporating the teaching sequence of contrast, generalization and fusion, students should be capable of easily grasping the knowledge of probability (Lo, 2012).

METHODS

This study examined the effectiveness of student learning in probability with the use of course materials specially designed under the framework of Variation Theory. A pretest-training-posttest design was conducted. The transfer performance, which was the improvement from the pretest to posttest scores, was considered to mirror the learning effectiveness of student participates. The performance of an experimental group (n = 25) enrolling in Secondary 5 in a school in Hong Kong was contrasted with that of two control groups (n = 66). One of the control groups was at similar academic ability (n = 30) as the experimental group, while the other group had higher academic ability (n = 36) according to the school streaming policy. In addition, for comparison purpose, a group of secondary 1 students (n = 32) was invited to write the same assessment at the time before the secondary 5 students started their learning of probability.

The assessment instrument used in this study was the excerpt of Garfield's (2003) Statistical Reasoning Assessment. As this study focuses on probability, the questions in Garfield's (2003) Assessment that taps on statistics were excluded. As a result, eleven questions were excerpted to focus on five areas in probability. The first two questions attempted to assess students' interpretation of probabilities; and the remaining 9 questions touched on students' computation of probability, among which Question 3 required the basic expression of probabilities as ratios, Questions 4 – 6 tested students on independence, Question 7 taped on sampling variability, and Questions 8 – 11 required

combinatorial reasoning (Appendix). The instrument also measured four probabilistic misconceptions, namely outcome orientation, representativeness, law of small number, and equiprobability bias. All participating students in both the experimental and control groups responded to the assessment immediately ahead of and subsequent to the 8 lessons of probability.

The hypothesis was that the understanding of probability acquired by the experimental group who used the teaching materials designed under Variation Theory was better than that by the control groups who underwent the traditional teaching approach without the framework of Variation Theory.

The study employed a causal-comparative design because the intact classes experienced the teaching intervention. It is widespread that educational programmes are delivered to the pre-existing classrooms (Maxey, 2013; Burke, 2000). Yet, this study could still control some extraneous variables, as the school streamed her students in accordance with their academic performance in the previous years and gender distribution. Furthermore, all three groups of students had the same amount of mathematics lesson time, a total of eight 65-minute lessons of probability, and were taught by teachers with similar 15 years of teaching experience.

RESULTS

Comparison on Probabilistic Skills between Secondary 1 Students and Secondary 5 Students Prior to Teaching Intervention

Although secondary 5 students had some basic training of probability two years ago when they were in secondary 3, their initial understanding on probabilistic skills did not outperform their fellow secondary 1 school mates. The original mean score of all secondary 5 students (0.44) prior to this study was close, or even slightly lower than, to that of their secondary 1 schoolmates (0.45). Among the three secondary 5 groups, the higher ability class (0.55) scored higher than the other two average ability classes (0.37 for the experimental group and 0.44 for the control group) in the overall probabilistic mean (Table 1).

Testing of Assumptions

Levene's test was performed for testing the homogeneity of variances in different probabilistic items and skills. Result showed that the assumption of homogeneous variances among four classes was not held in some skills, namely the skills on interpreting probabilities (Skill I) (p-value 0.026), understanding probabilistic as ratios (Skill IIa) (p-value 0.004), understanding of independence (Skill III) (p-value 0.084), and understanding sampling variability (Skill IV) (p-value 0.000), since their corresponding levels of significance in Levene's test were under 0.1, or even 0.05. In addition, Shapiro Wilk's test was conducted to test for the normality of the probabilistic skill scores. The significances for all probabilistic skills and individual items, except the overall aggregate scores, were below 0.05, indicating that the assumption of normal distribution for the individual probabilistic data was not hold. The parametric analysis was further investigated solely for the aggregated means, while the non-parametric ANOVA analysis, namely Kruskal Wallis Test, which was not sensitive to the violation of underlying assumptions, was used to further examine the initial mean score differences among the four classes.

Comparison between Levels

Secondary 5 students, though they had learnt more mathematics than secondary 1 students, did not necessarily outperform the younger counterparts in probabilistic knowledge. The overall mean scores on probabilistic skills for secondary 5 students were 0.44, while that for secondary 1 student was 0.45. There was no significant difference between secondary 1 and secondary 5 students before the topic of probability was taught (p-value = 0.871). To plough further into the three secondary 5 classes, the ANOVA showed the absence of a significant difference on the aggregate means among all four classes (F statistics = 0.362, p-value = 0.053). The higher ability secondary 5 class (0.50) scored higher than all its counterparts (0.45 for secondary 1 class, 0.44 for the average ability secondary 5 control class, and 0.37 for the secondary 5 experimental class). The discrepancy between the high ability control class and the experimental class was the most remarkable (t statistics = -2.907, p-value = 0.005).

Further non-parametric comparison was carried out on the subcategories between the senior and junior levels. Yet, contrast to our expectation, secondary 5 students who had learnt more proportions and ratios, along with some basic probabilities, than secondary 1 students did not necessarily score higher in all probabilistic skill scores. Among the four probabilistic skills measured in this study, the secondary 5 students could only score higher than the secondary 1 students in two categories, namely the score for computing probabilities (Skill II) (0.37 for secondary 5; 0.34 for secondary 1) and that for understanding independence (Skill III) (0.57 for secondary 5; 0.56 for secondary 1). The results of the Mann Whitney U test for the two independent samples concluded no significant discrepancies on all probabilistic skills between the secondary 1 and secondary 5 students.

Comparison among Classes

When the three senior secondary 5 students were treated individually, obvious discrepancies were detected among the classes in the scorings for understanding of independence (Skill III) and sampling variability (Skill IV). Kruskal Wallis test, which was a one-way ANOVA on ranks, showed that the score distributions for understanding independence (Skill III) (Chi-square statistics = 16.872; p-value = 0.001) and that for understanding sampling variability (Skill IV) (Chi-square statistics = 8.792; p-value = 0.032) were quite different among the four classes. The multiple pair wise comparison of Mann Whitney U tests with Bonferroni correction indicated that the score discrepancies mainly due to the difference from the high ability secondary 5 class. Significant score differences were detected on understanding independence (Skill III) between the high ability control class and the experimental class (p-value = 0.000), together with between the high and average ability control classes (p-value = 0.001). Similarly, there was marginally significant score difference in the scores for understanding sampling variability (Skill IV) between the average and the high ability control secondary 5 classes (p-value = 0.010).

Table 1: Result on the Probabilistic Skills with the Focus of the Mean Scores of Secondary 1 and the Initial Mean Scores of Secondary 5

	S1	S5								
	Comparison	All			Experimental	Control - Average Ability	Control - High Ability			
	n = 32	n = 91			n = 25	n = 30	n = 36			
	Mean Score	Pre Mean Score	Mann Whitney (S1 & S5) Sig.	Kruskal Wallis (all S5 classes) Sig.	Pre Mean Score	Pre Mean Score	Mann Whitney (Exp& Average Ability Control) Sig.	Pre Mean Score	Mann Whitney (Exp& High Ability Control) Sig.	Mann Whitney (High & Average Control) Sig.
Overall mean	0.45	0.44	0.945	0.051	0.37	0.44	0.111	0.50	0.005	0.233
I. For interpreting probabilities	0.63	0.52	0.187	0.387	0.46	0.50	0.697	0.57	0.304	0.467
Q1d	0.56	0.47	0.351	0.737	0.42	0.46	0.733	0.50	0.530	0.778
Q2d	0.69	0.57	0.240	0.439	0.50	0.54	0.799	0.64	0.289	0.408
II. For computing probabilities	0.34	0.37	0.277	0.653	0.35	0.36	0.665	0.39	0.484	0.716
IIa. For computing probabilities – Understanding probabilities as ratios	0.71	0.73	0.851	0.173	0.79	0.57	0.095	0.81	0.896	0.044
Q3c	0.71	0.73	0.851	0.173	0.79	0.57	0.095	0.81	0.896	0.044
IIb. for correctly computing probability – Using combinatorial reasoning	0.25	0.28	0.502	0.673	0.24	0.31	0.248	0.29	0.640	0.596
Q8a	0.28	0.36	0.402	0.824	0.33	0.39	0.660	0.36	0.827	0.796
Q9b	0.28	0.27	0.927	0.840	0.29	0.32	0.818	0.22	0.546	0.376
Q10a	0.16	0.18	0.745	0.414	0.08	0.18	0.321	0.25	0.105	0.496
Q11b	0.28	0.32	0.700	0.825	0.25	0.36	0.409	0.33	0.494	0.844
III. For understanding independence	0.56	0.57	0.901	0.001	0.42	0.50	0.227	0.73	0.000	0.001
Q4e	0.69	0.81	0.168	0.009	0.58	0.86	0.028	0.92	0.002	0.453
Q5df	0.50	0.33	0.106	0.008	0.29	0.14	0.195	0.51	0.092	0.002
Q6e	0.47	0.57	0.336	0.021	0.38	0.50	0.370	0.75	0.004	0.040
IV. For understanding sampling variability	0.32	0.26	0.514	0.032	0.17	0.46	0.024	0.17	1.000	0.010
Q7b	0.32	0.26	0.514	0.032	0.17	0.46	0.024	0.17	1.000	0.010

^ The score ranges from 0 to 1. The higher the score is, the better the understanding of probability students have.

Comparison on the Probabilistic Understanding of Secondary 5 Students Before and After the Teaching

Aspect 1: Probabilistic Skills

The experimental group could improve their probabilistic skills as their posttest mean score increased from the initial pretest by 24%. In contrast, the control group, which was at similar ability as the experimental group, showed no progress in terms of their transfer performance. The high ability control group at the same time could increase their posttest mean score by 12%. Among the four measured probabilistic skills in this study, the experimental group made the most improving transfer performance in two skills, namely computing probabilities (Skill II) and understanding independence (Skill III), as compared to two other control groups.

Testing of Assumptions

According to Levene's test results, the transfer scores measuring for the probabilistic skills or items acquired by the three secondary 5 groups basically fulfilled the assumption of equal variances, except for three individual items, namely one item from score for understanding probabilities as ratios (Skill IIa) and two items from score for understanding independence (Skill III).

Besides, Shapiro Wilk test, which was conducted for testing the data normality, found that at least one of the three secondary classes violated the normality assumption in any probabilistic skills or items. For instance, in terms of the

aggregated mean scores, the data acquired from the high ability control group could not fulfill the normality supposition, since the level of significance of the Shapiro Wilk test was 0.002, below 0.05, while the data acquired from the experimental group and the average ability control group showed that there was no violations of the postulation of normality as the corresponding levels of significance were over 0.05. Because the assumptions of equal variances and normality could not be simultaneously fulfilled, the probabilistic skill scores acquired by the three secondary 5 classes were further analyzed and compared through non-parametric analysis.

Comparison among Classes

Based on the comparison between the pretest and posttest scores of secondary 5 students, an obvious progress in transfer performance was found in the understanding of probability in the experimental group. The overall assessment mean score of the experimental class improved by 24%, from the pretest score of 0.37 to the posttest score of 0.46 (out of 1) with the p-value of 0.071 in the non-parametric Wilcoxon rank sum test, reflecting that the improvement made by the experimental group of students showed a statistical significance under the 90% confidence (Table 2).

The aggregate transfer improvement made by the experimental group was the highest as compared to the other two control groups. The increase percent of the experimental group was slightly higher than that of the high ability control group, even though the difference between these two groups was not significant according to the independent Mann-Whitney U test (p-value = 0.637). The capable control class could also improve remarkably with a 12% increase from the mean pretest score of 0.50 to the mean posttest score of 0.56, resulting in the p-value of 0.048 under the Wilcoxon rank-sum test. In contrast, the development of the probabilistic knowledge in the average ability control group, which was taught by the same teacher as the capable control class, was not notable as their mean score slightly reduced from 0.44 to 0.42, with a 5% drop, after the lessons. Out of the 11 individual probabilistic items, the experimental group and the high ability control group were able to score more marks in a total of 10 and 9 items respectively in the assessment after the training; yet, the control group of similar ability as the experiment group could advance in 5 items only.

Out of the four measured probabilistic skills, the experimental group could improve in three skills from the pretest to posttest assessment, whereas the average ability control group could advance in only one skill. The transfer performance achieved by the experimental group in the scores for computing probabilities (Skill II) and understanding independence (Skill III) was the highest among the three classes.

The experimental class progressed in particular well in the understanding of independence (Skill III), from the mean score 0.42 to 0.55 between the pretest and posttest with an increase of 31%. The advancement by the experimental group in this aspect was significantly higher than that by the control group of similar ability, according to the Mann-Whitney U test result under the 90% significance (p-value = 0.056). Under this category for understanding independence, which included three individual items, the experimental group could considerably improve its mean correct percent from 58% to 80% in one of the items which regarded flipping a fair coin 5 times, and such improvement was statistically significantly at the p-value of 0.034 in the Wilcoxon signed rank test; whereas the control group of similar ability encountered a drop in their correct percent between the pretest and the posttest performance from 86% to 68%. The difference between the experimental and the average ability control groups contributed to the significance of p-value 0.024 under the Mann Whitney rank sum test for the comparison of two independent samples, reflecting the existence of a strong statistically difference between the two group results.

The transfer performance of the experimental group also surpassed those of her counterparts in computing probabilities (Skill II). While the experimental group gained 26% from pretest score 0.35 to posttest score 0.44, the control groups of similar ability and higher ability achieved 11% and 15% respectively. Similar result was found in one of the individual items for measuring probabilities using the combinatorial reasoning (Skill IIb). The transfer correct percent in the item (Item 8) about rolling dice with 5 black faces achieved by the experimental group, with an improvement of 58%, was remarkably higher than that by the control group of similar ability, with a reduction of 26% in transfer performance, alongside also higher than the control group of high ability which had a reduction of 6%, ending with the significant class differences at the 90% confidence level under the Mann-Whiney rank-sum test. The posttest correct percent made by the experimental group in this item (0.52) went beyond the posttest correct percent of the average ability control group (0.29) and the higher ability control group (0.34).

Nevertheless the average capable students, no matter which experimental or control group they were in, in general failed to improve their abilities to interpret probabilities (Skill I) even after the lessons. The posttest mean scores for interpreting probabilities remained the same for the experimental group (0.46) and even decreased for the average capable control group (from 0.50 to 0.38 with a drop of 24%).

Regarding the understanding of sampling variability (Skill IV), both the experimental and the higher capable students improved their knowledge. The experimental group of students increased their transfer scores by 65% from 0.17 to 0.28. The improvement by the higher capable students in their understanding on sampling variability was even more remarkable, with the mean correct percent from 0.17 to 0.31 resulting in the p-value of 0.058 under the Wilcoxon rank-sum test for the paired samples. However, the control group with similar ability as the experimental group encountered a slight drop (of 7% from 0.46 to 0.43) in their transfer performance.

Table 2: Results on the Transfer Performance of Probabilistic Skills by Secondary 5 Students

	Experimental Group (n = 25)				Average Ability Control Group (n = 30)					Higher Ability Control Group (n = 36)				
	Pre	Post	% Change	Wilco -xon (Pre & Post) Sig.	Pre	Post	% Change	Wilco -xon (Pre & Post) Sig.	Mann-Whitney (Exp & Average Control) Sig.	Pre	Post	% Change	Wilco -xon (Pre & Post) Sig.	Mann-Whitney (Exp & High Control) Sig.
Overall mean (s.d.)	0.37 (0.18)	0.46 (0.15)	24%	0.071	0.44 (0.17)	0.42 (0.16)	-5%	0.856	0.125	0.50 (0.15)	0.56 (0.16)	12%	0.048	0.637
I. for interpreting probabilities	0.46	0.46	0%	0.527	0.50	0.38	-24%	0.320	0.727	0.57	0.64	12%	0.290	0.233
Q1d	0.42	0.58	38%	0.317	0.46	0.39	-15%	1.000	0.479	0.50	0.63	26%	0.248	0.878
Q2d	0.50	0.36	-28%	0.102	0.54	0.36	-33%	0.166	0.883	0.64	0.66	3%	0.782	0.182
II. for computing probabilities	0.35	0.44	26%	0.238	0.36	0.40	11%	0.244	0.895	0.39	0.45	15%	0.252	0.910
II. for computing probabilities – Understanding probabilities as ratios	0.79	0.88	11%	0.564	0.57	0.61	7%	0.739	0.978	0.81	0.86	6%	0.564	0.825
Q3c	0.79	0.88	11%	0.564	0.57	0.61	7%	0.739	0.978	0.81	0.86	6%	0.564	0.825
IIb. for computing probability – Using combinatorial reasoning	0.24	0.32	33%	0.241	0.31	0.35	13%	0.409	0.911	0.29	0.35	21%	0.300	0.945
Q8a	0.33	0.52	58%	0.058	0.39	0.29	-26%	0.593	0.090	0.36	0.34	-6%	1.000	0.073
Q9b	0.29	0.20	-31%	1.000	0.32	0.29	-9%	0.739	0.796	0.22	0.31	41%	0.206	0.396
Q10a	0.08	0.25	213%	0.257	0.18	0.36	100%	0.132	0.788	0.25	0.31	24%	0.405	0.712
Q11b	0.25	0.33	32%	1.000	0.36	0.46	28%	0.248	0.427	0.33	0.43	30%	0.285	0.523
III. for understanding independence	0.42	0.55	31%	0.103	0.50	0.47	-6%	0.641	0.056	0.73	0.78	7%	0.505	0.187
Q4e	0.58	0.80	38%	0.034	0.86	0.68	-21%	0.248	0.024	0.92	1.00	9%	0.083	0.066
Q5df	0.29	0.24	-17%	1.000	0.14	0.32	129%	0.132	0.258	0.51	0.49	-4%	1.000	1.000
Q6e	0.38	0.60	58%	0.166	0.50	0.41	-18%	0.317	0.077	0.75	0.86	15%	0.366	0.321
IV. Score for understanding sampling variability	0.17	0.28	65%	0.705	0.46	0.43	-7%	0.782	0.656	0.17	0.31	82%	0.058	0.432
Q7b	0.17	0.28	65%	0.705	0.46	0.43	-7%	0.782	0.656	0.17	0.31	82%	0.058	0.432

Note. The score ranges from 0 to 1. The higher the score is, the better the understanding of probability students have.

Aspect 2: Probabilistic Misconceptions

The assessment instrument also extracted students' performance on four probabilistic misconceptions. Regarding the elimination of the four measured misconceptions examined in this study, the experimental class could slightly, though not significantly, reduce their misconception in two aspects of misconceptions, the outcome orientation and law of small number, based on the comparison between the pretest and posttest performance comparison. Yet, meanwhile, the control class of the similar ability showed no improvement on any of the measured misconceptions. The control group of higher ability could improve on eliminating the four measured probabilistic misconceptions.

Testing of Assumptions

The Levene's test results indicated that most of the transfer scores for the probabilistic misconceptions attained by the three secondary 5 groups were distributed in homogeneous variances, except for two individual items under the misconception of representation and one item from the equiprobability bias misconception. On the other hand, the Shapiro Wilk normality test results unveiled that the transfer performance in probabilistic misconceptions failed to meet the fulfillment of normal distribution for most misconception items, except the experimental group's aggregated misconception mean score, as well as the higher ability control group's three mean scores on the aggregated misconception, the outcome orientation misconception and the equiprobability bias misconception. Since none of the misconception categories could meet the simultaneous requirements of homogeneous variances and normal distribution for all three secondary 5 classes, the forthcoming inferential analysis for the comparisons among the three classes was carried out on the non-parametric tests.

Comparison among Classes

The extent to which the experimental class cleared her misconceptions was more than the control class of the similar ability. While the experimental class significantly avoided two misconception items, the control class of the similar ability could not show any obvious improvements in all items. For example, under the context of flipping a fair dice (i.e. Item 6), notably more students in the experimental group had cleared their outcome orientation misconception after the lessons, with 63% of the students having the misconception in the pretest to 24% at the posttest, leading to the p-value 0.011 under the Wilcoxon paired T-test which implied a statistically significance at the 95% confidence. Similarly, the percentage of students in the experimental group who encountered the misconception of equiprobability bias under the context of rolling a fair dice 6 times (i.e. Item 8) dropped from 50% in the pretest to 24% in the posttest, with the p-value 0.052 under the Wilcoxon paired T-test, which again suggested a statistical significance at the 90% confidence. It appeared that the experimental class could avoid more misconceptions than the control class after the learning of the probability unit.

Analogous to the results on the probabilistic skills, the higher capable control group turned up to handle the four measured misconceptions better than their counterparts. Fewer of these higher achieving students had misconceptions in all these aspects (i.e. outcome orientation, representativeness, law of small number and equiprobability bias) in the posttest than the pretest, even though none of the drops was statistically significant.

CONCLUSIONS

Variation Theory could help students learn probability in an efficient way, especially for those with average academic ability. It was evident that the higher capable students mastered the probabilistic knowledge well and avoid more probabilistic misconceptions regardless of the teaching approach they received. In contrast, the average students learnt

probability more effectively if they were exposed to different metacognitive experiences, such as the structured organization of a schematic model and a solution strategy. When students were guided through the cognitive process under the framework of Variation Theory, student learning of probability could be enriched.

In future research, researchers may extend the sample group to include an experimental group of higher ability, in order to have a more comprehensive comparison across students with different abilities and examine the impact of variation on the learning of probability.

REFERENCES

1. *Ang, L. H. &Shahrill, M. (2014). Identifying students' specific misconceptions in learning probability. International Journal of Probability and Statistics, 3 (2), 23-29.*
2. *Batanero, C. &D´ıaz, C. (2012). Training school teachers to teach probability: reflections and challenges. Chilean Journal of Statistics, 3 (1), 3 – 13. Retrieved from http://chjs.deuv.cl/Vol3N1/ChJS-03-01-01.pdf.*
3. *Burke, J. (2000). It's (beyond) Time To Drop the Terms Causal-Comparative and Correlational Research in Educational Research Methods Textbooks. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA, April 24-28, 2000. Retrieved from http://files.eric.ed.gov/fulltext/ED445010.pdf.*
4. *Cheng, E. (2016). Learning through the Variation Theory: A Case Study. International Journal of Teaching and Learning in Higher Education. 28(2), 283-292.*
5. *Ekawati, R. & Lin, F. (2014). Designing Teacher Professional Development for Mathematics Teaching with Variation Theory. Indonesian Mathematical Society Journal on Mathematics Education, 5(2), 127-137.*
6. *Garfield, J. (2003). Assessing Statistical Reasoning. Statistics Education Research Journal, 2(1), 22-38.*
7. *Lam, H. C. (2013). On generalization and variation theory. Scandinavian Journal of Educational Research, 57(4), 343–356.*
8. *Li, J. (2000). Chinese students' understanding of probability. Unpublished doctoral thesis. Singapore: Nanyang Technological University.*
9. *Lo, M. L. (2012). Variation theory and the improvement of teaching and learning. Göteborg :ActaUniversitatisGothoburgensis. Retrieved from https://gupea.ub.gu.se/bitstream/2077/29645/5/gupea_2077_29645_5.pdf.*
10. *Maxey, K. (2013). Differentiated Instruction: Effects on Primary Students' Mathematics Achievement. Unpublished doctoral thesis. Prescott Valley, Arizona: Faculty of School of Education, Northcentral University.*
11. *Olteanu, C. &Olteanu, L. (2013). Enhancing mathematics communication using critical aspects and dimensions of variation. International Journal of Mathematical Education in Science and Technology, 44 (4), 513-522.*
12. *Rakes, C. R. (2010). Misconceptions in rational numbers, probability, algebra, and geometry. Doctoral dissertation, University of Louisville. Retrieved from http://digital.library.louisville.edu/utils/getfile/collection/etd/id/957/filename/958.pdf.*

APPENDIX

The excerpted Statistical Reasoning Assessment (Garfield, 2003, p. 33)

(Skill I. Correctly interprets probabilities)

1. The following message is printed on a bottle of prescription medication:

   WARNING: For applications to skin areas there is a 15% chance of developing a rash. If a rash develops, consult

your doctor

A. Don’t use the medication on your skin. There is a good chance of developing a rash.

B. For application to the skin, apply only 15% of the recommended dose.

C. If a rash develops, it will probably involve only 15% of the skin.

D. About 15 of 100 people who use this medication develop a rash.*

E. There is hardly a chance of getting a rash using this medication.

2. The observatory center wanted to decide the accuracy of their weather forecasts. They searched their records for those days when the forecaster had reported a 70% chance of rain. They compared these forecasts to records of whether or not it actually rained on those particular days. The forecast of 70% chance of rain can be considered very accurate if it rained on:

A. 95% - 100% of those days.

B. 85% - 94% of those days.

C. 75% - 84% of those days.

D. 65% - 74% of those days.*

E. 55% - 64% of those days.

3. Two containers, labeled A and B, are filled with red and blue marbles in the following quantities:

Container	Red	Blue
AB	660	440

Each container is shaken vigorously. After choosing one of the containers, you will reach in and, without looking, draw out a marble. If the marble is blue, you win $50. Which container gives you the best chance of drawing a blue marble?

A. Container A (with 6 red and 4 blue)

B. Container B (with 60 red and 40 blue)

C. Equal chances from each container*

(Skill II. Understands independence)

4. The following shows the results from flipping a fair coin 5 times. Which of the following sequences is most likely to result?

A. HHHTT

B. THHTH

C. THTTT

D. HTHTH

E. All four sequences are equally likely*

5. Select one or more explanations for the answer you gave for the item in Q4.

A. Since the coin is fair, you should get about equal numbers of heads and tails.

B. Since coin flipping is random, the coin should alternate frequently between landing heads and tails.

C. Any of the sequences could occur.

D. If you repeatedly flipped a coin five times, each of these sequences would occur about as often as any other sequence.*

E. If you get a couple of heads in a row, the probability of a tails on the next flip increases.

F. Every sequence of five flips has exactly the sample probability of occurring. *

6. The following shows the results from flipping a fair coin 5 times. Which of the following sequences is <u>least</u> likely to result?

A. HHHTT

B. THHTH

C. THTTT

D. HTHTH

E. All four sequences are equally unlikely*

(Skill III. Understands sampling variability)

7. Half of all newborns are girls and half are boys. Hospital A records an average of 50 births a day. Hospital B records an average of 10 births a day. On a particular day, which hospital is more likely to record 80% or more female births?

A. Hospital A (with 50 births a day)

B. Hospital B (with 10 births a day)*

C. The two hospitals are equally likely to record such an event.

(Skill IV. Correctly computes probability – Uses combinatorial reasoning)

8. Five faces of a fair dice are painted black, and one face is painted white. The dice is rolled six times. Which of the following results is more likely?

A. Black side up on five of the rolls; white side up on the other roll*

B. Black side up on all six rolls

C. A and B are equally likely.

9. When two dice are simultaneously thrown, the following two results may occur:

- Result 1: A 5 and a 6 are obtained.

- Result 2: A 5 is obtained twice.

Select the response that you agree most:

A. The chance of obtaining each of these results is equal.

B. There is more chance of obtaining Result 1.*

C. There is more chance of obtaining Result 2.

D. It is impossible to give an answer. Please explain why:

10. When three dice are simultaneously thrown, which of the following results is most likely to be obtained?

A. Result 1: A 5, a 3 and a 6*

B. Result 2: A 5 three times

C. Result 3: A 5 twice and a 3

D. All three results are equally likely.

11. When three dice are simultaneously thrown, which of the following results is least likely to be obtained?

A. Result 1: A 5, a 3 and a 6

B. Result 2: A 5 three times*

C. Result 3: A 5 twice and a 3

D. All three results are equally unlikely.

Note: The correct choice is marked with *.